On the gamma distribution with small shape parameter

Ryan Martin
Department of Mathematics, Statistics, and Computer Science
University of Illinois at Chicago
rgmartin@uic.edu

Chuanhai Liu
Department of Statistics
Purdue University
chuanhai@purdue.edu

February 11, 2013

Abstract

The gamma distribution with small shape parameter is difficult to characterize. In particular, standard algorithms for sampling from such a distribution often fail, so something special is needed. In this paper, we first obtain a limiting distribution for a suitably normalized gamma distribution when the shape parameter tends to zero. Then this limiting distribution provides insight to the construction of a new, simple, and highly efficient acceptance–rejection algorithm. Pseudo-code, and an R implementation, for this new sampling algorithm are provided.

Keywords and phrases: Acceptance–rejection method; asymptotic distribution; characteristic function; gamma function; R software.

1 Introduction

Let Y be a positive gamma distributed random variable with shape parameter $\alpha > 0$, denoted by $Y \sim \mathsf{Gam}(\alpha, 1)$. The probability density function for Y is given by

$$p_\alpha(y) = \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y}, \quad y > 0,$$

where the normalizing constant, $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} \, dy$, is the gamma function evaluated at α. This is an important distribution in statistics and probability modeling. In fact, since the gamma distribution is closely tied to so many important distributions, including normal, Poisson, exponential, chi-square, F, beta, and Dirichlet, one could argue that it is one of the most fundamental. But despite its importance, when the shape

parameter α is small, the gamma distribution is not so well understood. Consequently, seemingly routine calculations can become intractable. For example, in the popular R software, the functions related to the gamma distribution—in particular, the `rgamma` function for sampling—become relatively inaccurate when the shape parameter is small (R Development Core Team 2011, "GammaDist" documentation). To circumvent these difficulties, and to move towards new and more accurate and efficient software, an important question is if $\mathsf{Gam}(\alpha, 1)$, suitable normalized, has a meaningful non-degenerate limiting distribution as $\alpha \to 0$. In this paper, we give an affirmative answer to this question, and then we use the result to develop a new and improved algorithm—in terms of both accuracy and efficiency—for sampling from a small-shape gamma distribution.

When the shape parameter α is large, it follows from the infinite-divisibility of the gamma distribution and Lindeberg's central limit theorem (Billingsley 1995, Sec. 27) that the distribution of Y is approximately normal. Specifically, as $\alpha \to \infty$,

$$\alpha^{-1/2}(Y - \alpha) \to \mathsf{N}(0, 1) \quad \text{in distribution.}$$

For large finite α, better normal approximations can be obtained by working on different scales, such as $\log Y$ or $Y^{1/3}$. Our interest is in the opposite extreme case, where the shape parameter α is small, approaching zero. Here, neither infinite-divisibility nor the central limit theorem provide any help. In Theorem 1 below, we prove that $-\alpha \log Y$ converges in distribution to $\mathsf{Exp}(1)$, the unit-rate exponential distribution, as $\alpha \to 0$. Our proof is based on a basic property of the gamma function and Lévy's continuity theorem for characteristic functions.

Motivated by the limit distribution result in Theorem 1, we turn to the problem of simulating from a small-shape gamma distribution. This is a challenging problem with many proposed solutions; see Devroye (1986) and Tanizaki (2008). For small shape parameters, the default methods implemented in R and MATLAB, due to Ahrens and Dieter (1974) and Marsaglia and Tsang (2000), respectively, have some shortcomings in terms of accuracy and/or efficiency. The exponential limit in Theorem 1 for the normalized gamma distribution suggests a convenient and tight envelope function to be used in an acceptance–rejection sampler (Flury 1990). We flesh out the details of this new algorithm in Section 3 and provide an R function to implement this method online. This new method is simple and very efficient. Compared to the method implemented in `rgamma`, which frequently returns $\log Y = -\infty$ when $\alpha \leq 0.001$, ours can quickly easily produce genuine samples of $\log Y$ for even smaller α. It is also more efficient than the ratio-of-uniform samplers recently proposed in Xi et al. (2013). Our proposed strategy should also be useful for other problems.

2 Limit distribution result

We start with some notation. For the gamma function $\Gamma(z)$ defined above, write $f(z) = \log \Gamma(z)$. Then the digamma and trigamma functions are defined as $f_1(z) = f'(z)$ and $f_2(z) = f''(z)$, the first and second derivatives of the log gamma function $f(z)$. Recall that these are related to the mean and variance of $\log Y$, with $Y \sim \mathsf{Gam}(\alpha, 1)$:

$$\mathsf{E}_\alpha(\log Y) = f_1(\alpha) \quad \text{and} \quad \mathsf{V}_\alpha(\log Y) = f_2(\alpha).$$

These formulae are most directly seen by applying those well-known formulae for means and variances in regular exponential families (Brown 1986, Corollary 2.3). Next, write $Z = -\alpha \log Y$. To get some intuition for why multiplication by α is the right normalization, consider the following recurrence relations for the digamma and trigamma functions (Abramowitz and Stegun 1966, Chap. 6):

$$f_1(\alpha) = f_1(\alpha + 1) - 1/\alpha \quad \text{and} \quad f_2(\alpha) = f_2(\alpha + 1) + 1/\alpha^2.$$

Then, as $\alpha \to 0$,

$$\begin{aligned} \mathsf{E}_\alpha(Z) &= -\alpha f_1(\alpha) = -\alpha f_1(\alpha + 1) + 1 = O(1), \\ \mathsf{V}_\alpha(Z) &= \alpha^2 f_2(\alpha) = \alpha^2 f_2(\alpha + 1) + 1 = O(1). \end{aligned}$$

That is, multiplication by α stabilizes the first and second moments of $\log Y$. Towards a formal look at the limiting distribution of Z, define the characteristic function

$$\varphi_\alpha(t) = \mathsf{E}_\alpha(e^{itZ}) = \mathsf{E}_\alpha(Y^{-i\alpha t}) = \Gamma(\alpha - i\alpha t)/\Gamma(\alpha), \tag{1}$$

where $i = \sqrt{-1}$ is the complex unit.

Theorem 1. *For $Y \sim \mathsf{Gam}(\alpha, 1)$, $-\alpha \log Y \to \mathsf{Exp}(1)$ in distribution as $\alpha \to 0$.*

Proof. Set $Z = -\alpha \log Y$. The gamma function satisfies $\Gamma(z) = \Gamma(z + 1)/z$, so the characteristic function $\varphi_\alpha(t)$ for Z in (1) can be re-expressed as

$$\varphi_\alpha(t) = \frac{\Gamma(\alpha - i\alpha t)}{\Gamma(\alpha)} = \frac{\Gamma(1 + \alpha - i\alpha t)/(\alpha - i\alpha t)}{\Gamma(1 + \alpha)/\alpha} = \frac{1}{1 - it} \frac{\Gamma(1 + o_\alpha)}{\Gamma(1 + o_\alpha)},$$

where o_α are terms that vanish as $\alpha \to 0$. Since the gamma function is continuous at 1, the limit of $\varphi_\alpha(t)$ as $\alpha \to 0$ exists and is given by $1/(1 - it)$. This limit is exactly the characteristic function of $\mathsf{Exp}(1)$, so the claim follows by Lévy's continuity theorem. □

3 Simulating small-shape gamma variates

Simulating random variables is an important problem in statistics and other quantitative sciences. For example, simulation studies are often used to compare performance of competing statistical methods, and the Bayesian approach to statistical inference relies heavily on simulations since the posterior distributions on moderate- to high-dimensional parameter spaces rarely have a tractable closed-form expression. Here we shall demonstrate that the limiting distribution result in Theorem 1 helps provide an improved algorithm for simulating gamma random variables with small shape parameter, compared to that of Ahrens and Dieter (1974) implemented in the `rgamma` function in R.

For $Y \sim \mathsf{Gam}(\alpha, 1)$ with α near zero, let $Z = -\alpha \log Z$. To simulate from the distribution of Z, one might consider an acceptance–rejection scheme; see, for example, Lange (1999, Chap. 20.4) or Givens and Hoeting (2005, Chap. 6.2.3). For this, one needs an envelop function that bounds the target density and, when properly normalized, corresponds to the density function of a distribution that is easy to simulate from. By Theorem 1 we know that Z is approximately $\mathsf{Exp}(1)$ for $\alpha \approx 0$. More precisely, the

density $h_\alpha(z)$ of Z has a shape like e^{-z} for $z \geq 0$. Therefore, we expect that a function proportional to an $\mathsf{Exp}(1)$ density will provide a tight upper bound on $h_\alpha(z)$ for $z \geq 0$. We shall similarly try to bound $h_\alpha(z)$ by an oppositely-oriented exponential-type density for $z < 0$, as is standard in such problems.

The particular bounding envelop function $\eta_\alpha(z)$ is chosen to be as tight an upper bound as possible. This is done by picking optimal points of tangency with $h_\alpha(z)$. For this, we shall need a formula for $h_\alpha(z)$, up to norming constant, which is easily found:

$$h_\alpha(z) = ce^{-z-e^{-z/\alpha}}, \quad z \in (-\infty, \infty).$$

The norming constant c satisfies $c^{-1} = \Gamma(\alpha + 1)$. By following standard techniques, as described in Lange (1999, Chap. 20.4), we obtain the optimal envelope function

$$\eta_\alpha(z) = \begin{cases} ce^{-z}, & \text{for } z \geq 0, \\ cw\lambda e^{\lambda z}, & \text{for } z < 0, \end{cases}$$

where $\lambda = \lambda(\alpha) = \alpha^{-1} - 1$ and $w = w(\alpha) = \alpha/e(1-\alpha)$. A plot of the (un-normalized) target density $h_\alpha(z)$ along with the optimal envelope $\eta_\alpha(z)$ is shown in the left panel of Figure 1 for $\alpha = 0.1$. The normalized envelope function $\eta_\alpha(z)$ corresponds to the density function of a mixture of two (oppositely-oriented) exponential distributions, i.e.,

$$\frac{1}{1+w}\mathsf{Exp}(1) + \frac{w}{1+w}\{-\mathsf{Exp}(\lambda)\},$$

which is easy to sample from using standard tools, such as `runif` in R.

Pseudo-code for a new program, named `rgamma.ss`, for simulating a sample of size n from a $s$mall-$s$hape gamma distribution, on the log scale, based on this acceptance–rejection scheme is presented in Algorithm 1. R code for this program is also available at the first author's website (`www.math.uic.edu/~rgmartin`). Though the R code is already fast, a proper implementation should be done in C, with inline functions, test sample recycling (e.g. Rubin 1976), and perhaps other optimizations.

The acceptance rate $r(\alpha)$ for the proposed method is

$$r(\alpha) = \{1 + w(\alpha)\}^{-1} = \Big\{1 + \frac{\alpha}{e(1-\alpha)}\Big\}^{-1}, \tag{2}$$

which is plotted in the right panel of Figure 1 as a function of α near zero. It is clear from the graph, and also from the approximation $r(\alpha) \approx 1 - \alpha/e$ for $\alpha \approx 0$, that the acceptance rate converges to 1 as $\alpha \to 0$. This indicates the high efficiency of the proposed acceptance–rejection method when α is small. This is to be expected based on Theorem 1: when $\alpha \approx 0$, Z is approximately $\mathsf{Exp}(1)$, so an algorithm that proposes $\mathsf{Exp}(1)$ samples with probability $1/(1+w) \approx 1$ will almost always accept the proposal.

One can easily experiment with the new function `rgamma.ss` and the function `rgamma` built in to R. As the R documentation for `rgamma` indicates, for $\alpha \leq 0.001$, one will frequently observe $Y \equiv 0$, corresponding to samples $\log Y \equiv -\infty$. So, obviously, that function is not suitable for small α values. On the other hand, the new function `rgamma.ss` can easily produce genuine samples of $\log Y$ for even smaller α values. Moreover, the code is very simple and the sampling is highly efficient.

Algorithm 1 – Pseudo-code for program `rgamma.ss` to simulate n independent samples of $\log Y$, with $Y \sim \mathsf{Gam}(\alpha, 1)$ and small shape parameter α.

set $\lambda \leftarrow \lambda(\alpha)$, $w \leftarrow w(\alpha)$, and $r \leftarrow r(\alpha)$ as in the text.
for $i = 1, \ldots, n$ **do**
    loop
        $U \leftarrow \mathsf{Unif}$ #Unif denotes the random number generator
        if $U \leq r$ **then**
            $z \leftarrow -\log(U/r)$
        else
            $z \leftarrow \log(\mathsf{Unif})/\lambda$
        end if
        if $h_\alpha(z)/\eta_\alpha(z) > \mathsf{Unif}$ **then**
            $Z_i \leftarrow z$
            break
        end if
    end loop
    $\log Y_i \leftarrow -Z_i/\alpha$
end for

Acknowledgments

This work is partially supported by the U.S. National Science Foundation, grants DMS–1007678, DMS–1208833, and DMS–1208841.

References

Abramowitz, M. and Stegun, I. A. (1966), *Handbook of mathematical functions, with formulas, graphs, and mathematical tables*, New York: Dover Publications Inc.

Ahrens, J. H. and Dieter, U. (1974), "Computer methods for sampling from gamma, beta, Poisson and binomial distributions," *Computing (Arch. Elektron. Rechnen)*, 12, 223–246.

Billingsley, P. (1995), *Probability and measure*, New York: John Wiley & Sons Inc., 3rd ed.

Brown, L. D. (1986), *Fundamentals of statistical exponential families with applications in statistical decision theory*, Institute of Mathematical Statistics Lecture Notes—Monograph Series, 9, Hayward, CA: Institute of Mathematical Statistics.

Devroye, L. (1986), *Nonuniform random variate generation*, New York: Springer-Verlag.

Flury, B. D. (1990), "Acceptance-rejection sampling made easy," *SIAM Rev.*, 32, 474–476.

Givens, G. H. and Hoeting, J. A. (2005), *Computational statistics*, Wiley Series in Probability and Statistics, Hoboken, NJ: Wiley-Interscience [John Wiley & Sons].

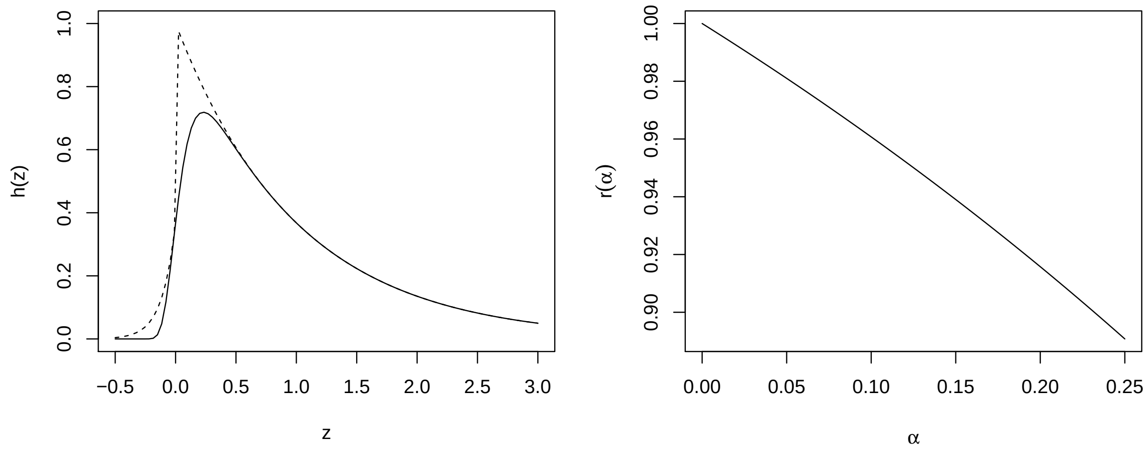

Figure 1: Left: the (un-normalized) $h_\alpha(z)$ and envelope $\eta_\alpha(z)$ for $\alpha = 0.1$; Right: the acceptance rate $r(\alpha)$ in (2) as a function of α.

Lange, K. (1999), *Numerical analysis for statisticians*, Statistics and Computing, New York: Springer-Verlag.

Marsaglia, G. and Tsang, W. W. (2000), "A simple method for generating gamma variables," *ACM Trans. Math. Software*, 26, 363–372.

R Development Core Team (2011), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Rubin, H. (1976), "Some fast methods of generating random variables with preassigned distributions. I. General acceptance–rejection procedures," Tech. Rep. 466, Purdue University, Department of Statistics, West Lafayette, IN.

Tanizaki, H. (2008), "A simple gamma random number generator for arbitrary shape parameters," *Econ. Bull.*, 3, 1–10.

Xi, B., Tan, K. M., and Liu, C. (2013), "Logarithmic transformation based gamma random number generators," *J. Statist. Softw.*, to appear.